The Evolution of Galaxies on Cosmological Timescales
ASP Conference Series, Vol. 3×10^8, *1999*
J. E. Beckman, and T. J. Mahoney, eds.

Large-Scale Visualization of Digital Sky Surveys

Joseph C. Jacob

Jet Propulsion Laboratory, M/S 126-234, 4800 Oak Grove Drive, Pasadena, CA 91109-8099

Laura E. Husman

Jet Propulsion Laboratory, M/S 126-234, 4800 Oak Grove Drive, Pasadena, CA 91109-8099

Abstract. Designed and developed within JPL/Caltech's supercomputing and visualization environment, the virtual observatory software described in this paper permits high performance visualization and analysis of large sky images and catalogs. Functionality includes parallel generation of large sky image mosaics and synthetic maps from catalogs, techniques for interactive exploration of large images on workstation or special high resolution displays, image enhancement features, catalog viewing, catalog to image relation, and run-time compositing of multi-resolution datasets. The tools and technologies being developed are intended to be components of the National (and eventually Global) Virtual Observatory effort, which will federate and provide seamless access to a wide variety of geographically distributed, multi-terabyte datasets.

1. Introduction

Modern sky surveys have archived many terabytes of catalog and image data at various resolutions and wavelengths. The great advances that have been made in remote sensing pose a tremendous challenge to the state of the art of information technology. Sheer volumes of scientific data are of little value without an effective ability to analyze them quickly and thoroughly, and to distill the essence of scientific knowledge from them.

The *Digital Sky* project (`http://www.digital-sky.org`) is an effort to design and prototype an interactive database search, analysis and visualization system spanning the catalog and image archives of multiple sky surveys. The work reported in this paper is the product of the *Digital Sky Virtual Observatory* task, which provides high-performance visualization, supercomputer level analysis (where appropriate), and large-scale data management for the *Digital Sky* project. The objective of this work is to prototype a "virtual observatory" that allows multi-spectral, multi-survey viewing and analysis of sky data, both images and catalogs, at the terabyte level.

The datasets that we are currently using to develop and test this new technology are at one arc minute resolution, Infrared Astronomical Satellite (IRAS)[1] in the infrared, and at one arc second, Digital Palomar Observatory Sky Survey (DPOSS)[2] in the visible and Two Micron All Sky Survey (2MASS)[3] in the near infrared. DPOSS and 2MASS each have multiple terabyte archives, making them ideal test surveys for development of technologies essential for a National Virtual Observatory (Boroson et al. 2000).

2. High Performance Information Technology Environment

The supercomputing and visualization environment at JPL and Caltech is well-suited for the large scale visualization and interactive analysis tool described in this paper. The hardware includes the following: SGI Onyx2 with 128 processors (peak speed of 76 Gflops), 32 GB of memory, and 6 InfiniteReality2 graphics pipes; HP V2500 with 128 processors, 128 GB of memory, and 1.8 TB of disk space; two 300 TB mass storage systems; 3×2 "PowerWall" display; and $4\times$ Gigabit ethernet and OC-12 communications interconnecting all major assets. The PowerWall display is a 3×2 matrix of 1280×1024 displays that are synchronized by the software to provide an effective display resolution of 3840×2048.

3. Large Image Mosaicking

The virtual observatory software that we are developing can be used both to generate large image mosaics from sky image patches and to view these large images. The mosaicking software is fully automated and can be run in parallel across 1,2,4, ...128 processors on the Onyx2. The input image patches may be in the common FITS format at any resolution, in any coordinate system and projection, and having any data type supported by FITS. The software translates these into the user-selected resolution and coordinate system of the output mosaic. The World Coordinate System library (Greisen & Calabretta 1999) is used to convert the input FITS images into a common output coordinate system and projection. The galactic, ecliptic, J2000, and B1950 coordinate systems are supported.

To date, the following sky mosaics have been constructed using this software: all-sky IRAS mosaic at 1 arc minute resolution, constructed from 430 individual image patches with 4 bands each; $10^\circ \times 10^\circ$ 2MASS mosaic at 1 arc second resolution constructed from over 1000 individual images with 3 bands each; $2^\circ \times 2^\circ$ DPOSS mosaic at 1 arc second resolution, constructed from 100 individual images with 3 bands each; $14^\circ \times 13^\circ$ DPOSS mosaic constructed from two full DPOSS plates. In addition, the 2MASS and smaller DPOSS mosaic

---

[1] `http://www.ipac.caltech.edu/ipac/iras/iras.html`

[2] `http://astro.caltech.edu/~rrg/science/dposs_public.html`

[3] `http://www.ipac.caltech.edu/2mass`

listed above were registered to each other and combined to produce a 6 band multi-spectral mosaic that includes both the visible and near infrared.

4. Large Image Exploration

4.1. Large-Scale Image Input

The size of a mosaic that can be constructed and viewed is limited only by the available disk space. A full sky mosaic at 1 arc second resolution in one spectral band approaches one terabyte in size. For high performance viewing of these large mosaics, the data is stored in a custom hierarchical, tiled format with two defining characteristics. The data is stored as a series of 512×512 pixel "tiles," and at multiple resolutions, including full resolution, half resolution, quarter resolution, all the way down to the resolution where the entire image fits in a single tile. The tiled nature of the data storage format permits any subset of the data to be quickly referenced and extracted for viewing without requiring that all of the data be read into memory. It also permits the image to be quickly extracted and served for viewing at any arbitrary resolution by resampling with a kernel no larger than 2×2 pixels.

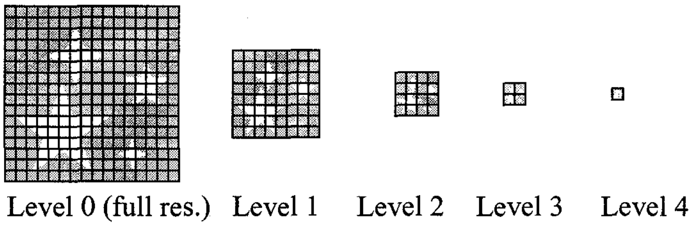

Figure 1. Hierarchical, tiled data format. Each level represents as 512×512 pixel tiles the entire image at a resolution which is half the resolution of the previous level.

In summary, the tiled nature of the data permits rapid panning to any arbitrary location and the hierarchical resolution storage permits rapid zooming to any arbitrary zoom level. Also, intelligent data caching keeps recently visited tiles in memory for rapid retrieval. In addition, the software includes a lookahead cache that loads additional tiles just outside the bounds of what is visible on the display and at neighboring resolution layers for improved performance. The disk storage penalty paid for this rapid panning and zooming capability is about one-third the size of the full resolution image.

4.2. Large-Scale Image Output

Just as the image viewing software scales on the input side, the software also scales on the output side, allowing displays ranging from single screen workstation displays up to large PowerWall displays (described in Section 2.). Single pipe and multi-pipe support is provided. The software can be configured to use any rectangular subset of the available display screens. The supported display configurations are specified in a configuration file and are easily selectable at run-time with a command line option or by setting an environment variable.

High performance on a 3×2 PowerWall display (3840×2048 effective resolution) has been demonstrated.

4.3. Image and Catalog Navigation and Viewing

The software that has been developed permits high performance visualization of both sky image and catalog data. A number of features are provided to enable efficient navigation of the potentially huge images. Mouse and keyboard interfaces for smooth, variable speed panning and zooming are provided. A *Global Map View* of the dataset is a popup window that shows the entire dataset with a box highlighting the region that is currently visible on the display. Users may jump to any location in the image by clicking at the location in the Global Map View. Another way to quickly jump to any location in the image is to specify that location in either pixel or sky coordinates in the *Coordinate Selection* window. Galactic, ecliptic, and celestial coordinate systems are supported. The Coordinate Selection window may also be used to retrieve the pixel or sky coordinate of any pixel that is selected by clicking with the mouse on the image.

The image mosaicking software described in Section 3. is capable of constructing mosaics having many spectral bands. The image viewing software permits any subset of three bands to be mapped to red, green, and blue at runtime. This allows the user to do such things as view both visible and infrared bands from different sky surveys simultaneously and then switch back to all visible or all infrared.

Keyboard and graphical user interfaces for brightness and contrast adjustments are provided, and may be applied to all three video channels (red, green, and blue), or to any channel individually. For example, this feature allows viewing of both Andromeda and the center of the Milky Way in our IRAS mosaic, as illustrated in Figure 2.

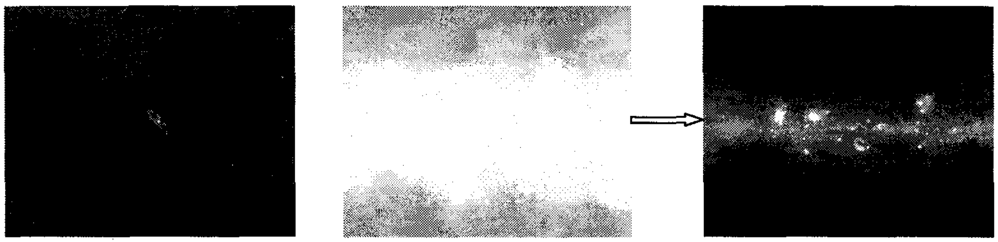

Figure 2. In the full sky IRAS image, the brightness and contrast settings for optimal viewing of the Andromeda galaxy (left) causes the center of the Milky Way to saturate (center). The brightness and contrast can then be adjusted to enhance the structure at the center of the Milky Way (right).

The catalogs may be viewed in ASCII text in a scrollable window and as image overlays, as illustrated in Figure 3, or as synthetic maps. With overlays which are vectors drawn over the image, the shape, size, and color of the overlay objects may be set according to the values in one or more columns of the catalog. With synthetic maps, the catalog positions and values are used to generated a pixelated image. For instance, the Hipparcos catalog was used to generate a synthetic map of the sky with 118,218 stars down to magnitude 14.

The image and catalog viewing capabilities are tightly coupled, allowing easy relation of a location in the images to catalog entries for celestial objects in the proximity and vice versa. For instance, the user may select a region of the sky in an image and see the catalog entries for those objects in that region highlighted in both the image and in the catalog window, as illustrated in Figure 3. Alternatively, the user may select a catalog entry in the scrollable list and see that object highlighted in the image or jump to the position of that object in the image. This demonstrates both image to catalog and catalog to image relations.

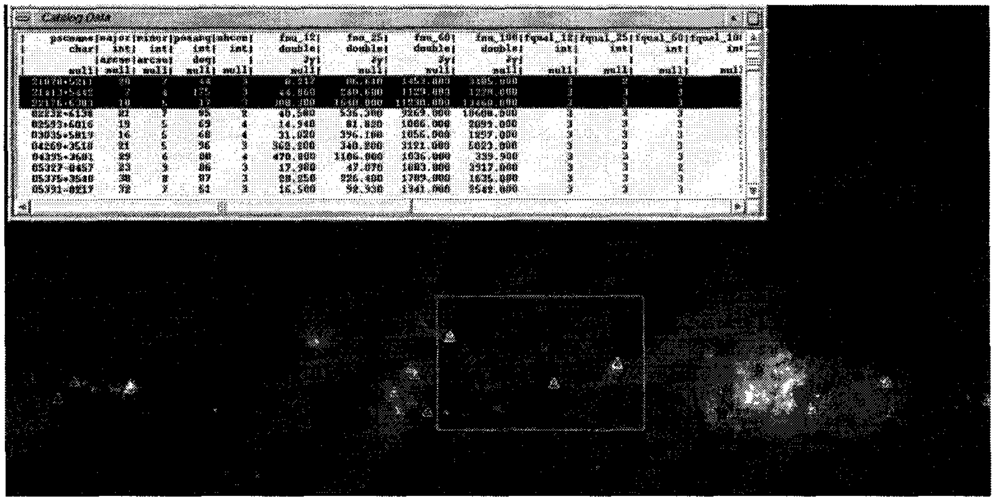

Figure 3. Image to catalog relation. The catalog shown here was obtained from a search of the IRAS catalog via the IRSA CatScan website `http://irsa.ipac.caltech.edu/applications/CatScan/`.

5. Automatic Dataset Compositing

The image viewing software described in this paper has support for fully automated **run-time** compositing of multiple datasets, correctly positioned based on pixel resolution and latitude and longitude at a corner. Any number of datasets may be composited, although panning and zooming performance is degraded slightly with each one added. This capability allows the user to do such things as view high resolution insets of particular celestial objects or regions of the sky overlaid on top of lower resolution imagery of the whole sky. As an example, refer to Figure 4 which shows four composited datasets, all-sky IRAS and Hipparcos synthetic map at 1 arc minute resolution, and high resolution insets of 2MASS and DPOSS at 1 arc second. The user can start at a zoom setting that permits synoptical viewing of the whole sky, smoothly zoom in to full resolution of IRAS, and seamlessly continue zooming in to the full resolution of the high resolution datasets.

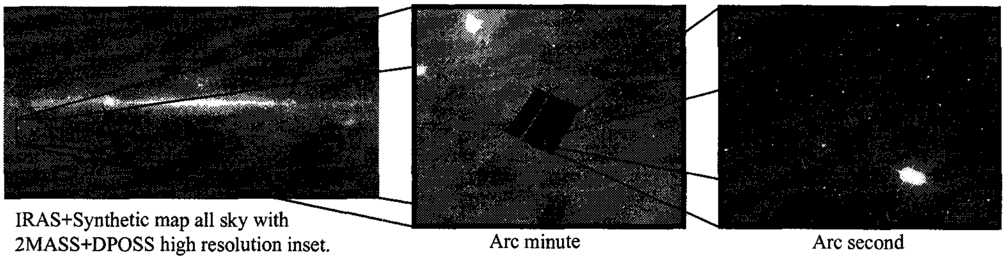

Figure 4. Automatic dataset compositing.

6. Conclusion

In this paper, the new technology we are developing for generation of large sky image mosaics and viewing of the large images and sky catalogs was described. Many capabilities are included for large image and catalog navigation and exploration, including both smooth and discontinuous pan and zoom over the whole sky and at any resolution. The catalogs may be viewed as ASCII tables, image overlays, or synthetic sky maps. Multiple image datasets of differing resolutions may be composited at run-time and viewed simultaneously, enabling novel multi-spectral views of the sky. The image and catalog viewing capabilities are tightly coupled, allowing easy relation from image to catalog and vice versa. The software is scalable both on the input side, permitting arbitrary sized images, and on the output side, permitting use of high resolution PowerWall displays.

Acknowledgments. We are grateful to our Digital Sky colleagues Robert Brunner, George Djorgovski, John Good, Ashish Mahabal, and Steve Odewahn. This work is sponsored by NASA Code SR, JPL Space and Earth Science Programs Directorate, Space Science Applications of Information Technology Program (formerly Interactive Analysis Environments Program).

References

Boroson, T., Brunner, R., De Young, D., Djorgovski, S., Hanisch, R., Strom, S., & Tody, D., 2000, White Paper, Toward a National Virtual Observatory: Science Goals, Technical Challenges, and Implementation Plan[4]

Greisen, E. W. & Calabretta, M., 1999, Representations of world coordinates in FITS, submitted to A&A[5]

[4] `http://astro.caltech.edu/nvoconf/white_paper.ps`

[5] `http://www.atnf.csiro.au/computing/software/wcslib.html`